

1. Estimateurs pour la régression linéaire simple (p.377)

Dans le cadre d'un problème de régression avec un prédicteur (c'est-à-dire $E(Y|X) = \beta_0 + \beta_1 X$), les estimateurs de β_0 et β_1 sont :

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} = \frac{(\sum_{i=1}^n x_i y_i) - n\bar{x}\bar{y}}{(\sum_{i=1}^n x_i^2) - n\bar{x}^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

2. Table d'ANOVA (p.384)

Soit la variable de réponse y avec les prédicteurs $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ sur un ensemble de n observations. Notez bien que la régression linéaire simple est le cas où $k = 1$. La table d'ANOVA du cas général est présenté dans le tableau ci-dessous. On appelle SST la variabilité totale des y_i

Source	Degré de liberté (dl)	Somme des carrés	Somme des carrés moyens	Statistique de Fisher	Valeur-p
Régression	k	SSR	$\frac{SSR}{k}$	$F = \frac{SSR/k}{SSE/(n-k-1)}$	$p(F_{k,n-k-1} > F)$
Erreur	n-k-1	SSE	$\frac{SSE}{n-k-1}$		
Total	n-1	SST			

Considérons le cas où $k = 1$. La décomposition suivante est possible :

$$\underbrace{\sum_{i=1}^n (y_i - \bar{y})^2}_{SST} = \underbrace{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}_{SSR} + \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{SSE}$$

Par le tableau précédant, on a que SST a $n - 1$ degrés de liberté, SSR et SSE ont respectivement 1 et $n - 2$ degrés de liberté. On a d'ailleurs les résultats suivants :

$$SSR = \hat{\beta}_1^2 s_{xx} = \hat{\beta}_1 s_{xy}, \quad \hat{\sigma}^2 = \frac{SSE}{n-2}, \quad F_0 = \frac{SSR/1}{SSE/(n-2)} = \frac{MSR}{MSE}, \quad R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

3. Intervalles de confiance (p.387)

Les intervalles suivants sont obtenus en considérant la régression linéaire simple.

L'intervalle de confiance pour l'ordonnée à l'origine β_1

Si les erreurs sont normales et indépendantes, alors un intervalle de confiance à $100(1 - \alpha)\%$ pour la pente β_1 est donné par :

$$\hat{\beta}_1 \pm t_{\frac{\alpha}{2}; n-2} \sqrt{\frac{MSE}{s_{xx}}}$$

L'intervalle de confiance pour la valeur moyenne au point x_0

On peut établir un intervalle de confiance à $100(1 - \alpha)\%$ pour la droite de régression au point $X = x_0$:

$$\hat{Y}_0 \pm t_{\frac{\alpha}{2}; n-2} \sqrt{MSE \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_{xx}} \right)}$$

L'intervalle de prévision pour une future observation au point x_0

On peut établir un intervalle de prévision à $100(1 - \alpha)\%$ pour une future observation au point x_0 :

$$\hat{y}_0 \pm t_{\frac{\alpha}{2}; n-2} \sqrt{MSE \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_{xx}} \right)}$$